

NBER WORKING PAPER SERIES

THE OPTIMAL DESIGN OF UNEMPLOYMENT
INSURANCE AND EMPLOYMENT PROTECTION.
A FIRST PASS

Olivier Blanchard
Jean Tirole

Working Paper 10443
<http://www.nber.org/papers/w10443>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2004

We are grateful to Daron Acemoglu, Suman Basu, Javier Ortega, John Hassler, and to the participants to the third Toulouse Macroeconomics workshop for helpful comments. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

©2004 by Olivier Blanchard and Jean Tirole. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Optimal Design of Unemployment Insurance and Employment Protection. A First Pass
Olivier Blanchard and Jean Tirole
NBER Working Paper No. 10443
April 2004, Revised November 2005
JEL No. D60, E62, H21, J30, J32, J38, J65

ABSTRACT

Much of the policy discussion of labor market institutions has been at the margin, with proposals to tighten unemployment benefits, reduce employment protection, and so on. There has been little discussion however of what the ultimate goal and architecture should be. The paper focuses on characterizing this ultimate goal, the optimal architecture of labor market institutions.

We start our analysis with a simple benchmark, with risk averse workers, risk neutral firms and random shocks to productivity. In this benchmark, we show that optimality requires both unemployment insurance and employment protection---in the form of layoff taxes; it also requires that layoff taxes be equal to unemployment benefits.

We then explore the implications of four broad categories of deviations: limits on insurance, limits on layoff taxes, ex-post wage bargaining, and heterogeneity of firms or workers. We show how the architecture must be modified in each case. The scope for insurance may be more limited than in the benchmark; so may the scope for employment protection. The general principle remains however, namely the need to look at unemployment insurance and employment protection together, rather than in isolation.

Olivier Blanchard
Department of Economics
MIT
Cambridge, MA 02139
and NBER
blanchar@mit.edu

Jean Tirole
IDEI
21 Allées de Brienne
31000 Toulouse
France
tirole@cict.fr

Introduction

Unemployment insurance and employment protection are typically discussed and studied in isolation. In this paper, we argue that they are tightly linked, and we focus on their optimal joint design.

We start our analysis in Section 1 with a simple benchmark. Workers are risk averse; entrepreneurs run firms and are risk neutral. The productivity of any worker-firm match is random. If productivity is low enough, the worker and the firm may separate, in which case the worker becomes unemployed.

In this benchmark, a simple way to achieve the optimum is for the state to pay unemployment benefits so as to insure workers, and to levy layoff taxes so as to lead firms to internalize the cost of unemployment and take an efficient layoff decision. Thus, the benchmark shows the tight conceptual relation between unemployment insurance and employment protection—defined as layoff taxes.

The optimum has two further characteristics: The first is that layoff taxes are equal to unemployment benefits: This common level delivers both full insurance and production efficiency. The second, which follows from the first, is that state intervention is not needed. The same allocation is achieved by having firms voluntarily pay severance payments to their workers; in effect, severance payments act both as unemployment insurance and layoff taxes.

This leads us to examine, in Sections 2 to 5, how these conclusions are affected by the introduction of four empirically-relevant deviations from the benchmark, from limits on unemployment insurance, to limits on the ability of firms to pay layoff taxes, to ex-post wage bargaining, and to ex-ante heterogeneity of either workers or firms. In each case, we ask two questions: The first is how the distortion affects the optimal combination of unemployment insurance and layoff taxes. The second is whether state intervention is required.

We find that, in general, efficiency and insurance require different levels of layoff taxes and unemployment benefits, with the difference being financed through payroll taxes — positive or negative.

The fact that layoff taxes are in general not equal to unemployment benefits also implies that individual firms can no longer implement the optimum on their own (i.e. through severance payments). A pooling or insurance agency is needed to receive payments both from firms which lay workers off and from firms that do not, and to distribute unemployment

benefits to laid off workers. We show that, in some cases, this role can be filled by a private agency, with voluntary participation by firms, but that in others, participation must be compulsory, implying a clear role for the state. We also show, in Section 6, that, while, in some cases, insurance can be provided through a combination of severance payments by firms and unemployment benefits by the agency, in other cases, the agency must require exclusivity in the provision of insurance, lest the outcome be suboptimal.

Before we start, we want to emphasize the limits of our analysis. Our framework is a static, one-period, model. As such, it represents a substantial step back from recent dynamic models of either unemployment insurance or employment protection. We do so for two reasons. First, we want to focus on the joint design of unemployment insurance and employment protection, which makes things more difficult. Second, we want to explore a large number of labor market imperfections, starting from as simple a benchmark as feasible. We believe that the conclusions we draw from our model, about the articulation of unemployment insurance and employment protection, about the relative use of layoff and payroll taxes in response to various market imperfections, and about the role of the state, are quite general and can help design better systems. In that context, we draw, in Section 7, what we see as our main conclusions and our intended extensions.

1 A benchmark

1.1 Assumptions

Tastes and technology are as follows:

- The economy is composed of a continuum of mass 1 of workers, a continuum of mass (at least) 1 of entrepreneurs, and the state.
- *Entrepreneurs* are risk neutral. Each entrepreneur can start and run a firm. There is a fixed cost of creating a firm, I , which is the same for all entrepreneurs.

If a firm is created, a worker is hired, and the productivity of the match is then revealed. Productivity is given by y from cdf $G(y)$, with density $g(y)$ on $[0, 1]$. The firm can either keep the worker and produce, or lay the worker off, who then becomes unemployed.

Realizations are iid across firms; there is no aggregate risk.

- The *firm*, but not the worker (or for that matter third parties such as an insurance

company or the state) observes y .¹

- *Workers* are risk averse, with utility function $U(\cdot)$. Absent unemployment benefits, utility if unemployed is given by $U(b)$ (so b is the wage equivalent of being unemployed).

1.2 The optimal allocation

Let \bar{y} be the threshold level of productivity below which workers are laid off. Let w be the payment to the workers who remain employed, and μ be the payment to the workers who are laid off.

The optimal allocation maximizes expected worker utility subject to the economy's resource constraint:²

$$\max_{\{w, \mu, \bar{y}\}} V_W \equiv G(\bar{y})U(b + \mu) + (1 - G(\bar{y}))U(w)$$

subject to:

$$V \equiv -G(\bar{y})\mu + \int_{\bar{y}}^1 y dG(y) - (1 - G(\bar{y}))w = I$$

From the first-order conditions, it follows that:

$$w^* = b + \mu^* \tag{1}$$

$$\bar{y}^* = b \tag{2}$$

Given \bar{y}^* , the levels of w^* and μ^* are determined by the resource constraint.

Condition (1) is an *insurance condition*: Workers achieve the same level of utility, whether employed or laid-off and unemployed.

Condition (2) is an *efficiency condition*: From the point of view of total output, it is efficient for firms to produce so long as productivity exceeds the wage equivalent of being unemployed (we shall call b the production-efficient threshold level).

¹This assumption is maintained throughout the paper, except in Section 4 which introduces ex-post bargaining, and in which both the worker and the firm observe the realization of y .

²We derive the optimal allocation ignoring the assumption that y is observed only by the firm. We shall show below that this optimal allocation can indeed be implemented.

1.3 Implementation

Any provision of insurance by the firm (a severance payment) can be duplicated by the state through unemployment benefits in this economy. At best, severance payments are redundant; in some circumstances (those envisioned in Section 2), we shall see that they must actually be ruled out in order to implement the optimal allocation. So, until we discuss alternative implementation schemes in Section 6, we assume away severance payments.

Consider now the following implementation of the optimal benchmark allocation:

- *Stage 1.* The state chooses a payroll tax rate τ , a layoff tax rate f , and unemployment benefits μ .

- *Stage 2.* Entrepreneurs decide whether to start firms and pay the fixed cost.

They offer contracts to workers. Contracts are characterized, explicitly, by a wage w , and, implicitly (since y is not contractable), a threshold productivity level \bar{y} below which the worker is laid off.

As all firms face the same cost and distribution of productivity, in equilibrium, all workers are initially hired.

- *Stage 3.* The productivity of each job is realized. Firms decide whether to keep or dismiss workers.

To show how the optimal allocation can be implemented, we work backwards in time.

At *Stage 3*, the cutoff \bar{y} is such that the firm is indifferent between keeping the worker and paying $w + \tau$ in wage and payroll tax and dismissing the worker and paying layoff tax f :

$$\bar{y} = w + \tau - f. \quad (3)$$

If $y > \bar{y}$, the firm keeps the worker, produces y , pays w to the worker, and τ to the state. If $y < \bar{y}$, the firm lays the worker off, pays f to the state; the state pays μ to the worker.

At *Stage 2*, firms' wage offer w satisfies the free entry condition:

$$V_F \equiv -G(\bar{y})f + \int_{\bar{y}}^1 y dG(y) - (1 - G(\bar{y}))(w + \tau) = I. \quad (4)$$

Consider now the *problem faced by the government* in choosing taxes and unemployment benefits at *Stage 1*. Condition (3) implies that to induce firms to take the production-efficient layoff decision $\bar{y}^* = b$, the following condition must hold:

$$w + \tau - f = b. \quad (5)$$

Because optimal insurance further requires that $w = b + \mu$, the state's policy must satisfy:

$$f - \tau = \mu. \quad (6)$$

The net fiscal cost to the firm of laying off a worker must be equal to the unemployment benefits paid to the worker by the state. Note that this condition implies a *positive* relation between the layoff and the payroll tax rates: For given unemployment benefits, the higher the payroll tax, the higher the layoff tax needed to induce the firm to take the production-efficient decision.

The government budget constraint implies a second relation between taxes and benefits:

$$V_G \equiv -G(\bar{y})(\mu - f) + (1 - G(\bar{y}))\tau = 0 \quad (7)$$

This constraint implies a *negative* relation between the layoff and the payroll tax rates: For given unemployment benefits, the higher the payroll tax, the lower the layoff tax required to balance the budget. Combining the two conditions gives:

$$f = \mu, \quad \tau = 0 \quad (8)$$

The layoff tax must be equal to unemployment benefits, and the payroll tax rate is equal to zero. The payment of layoff taxes equal to unemployment benefits makes firms fully internalize the cost of insurance provided by the state to the unemployed.

We summarize our results in Proposition 1.

Proposition 1. *In the benchmark, the optimal allocation is such that workers are fully insured ($b + \mu^* = w^*$), and the threshold productivity is equal to the wage equivalent of being unemployed ($\bar{y}^* = b$).*

Implementation is achieved through unemployment benefits equal to μ^ , and layoff taxes $f = \mu^*$. Payroll taxes are equal to zero. Put another way, the contribution rate, defined as the ratio of layoff taxes to unemployment benefits, is equal to one.*

2 Limits on Insurance

In our benchmark, workers could be and were fully insured. In practice, there are various reasons why this may not be feasible. Workers may require incentives not to shirk when employed, or to search when unemployed. Or there may be a non-pecuniary loss associated with becoming unemployed. We explore the implications of this last assumption, and return to a discussion of other potential reasons later.

Assume that the utility of workers is now given by $U(c)$ if employed, and by $U(c) - B$ if unemployed, so $B > 0$ is the utility cost of being unemployed. All other assumptions are the same as in the benchmark.

2.1 The optimal allocation

The optimal allocation is the solution to:³

$$\max_{\{w, \mu, \bar{y}\}} V_W \equiv G(\bar{y})(U(b + \mu) - B) + (1 - G(\bar{y}))U(w),$$

subject to the resource constraint:

$$V \equiv -G(\bar{y})\mu + \int_{\bar{y}}^1 y dG(y) - (1 - G(\bar{y}))w = I.$$

From the first-order conditions, it follows that:

$$w^* = b + \mu^*. \tag{9}$$

$$\bar{y}^* = b - \frac{B}{U'(w^*)}. \tag{10}$$

Given \bar{y}^* , the levels of w^* and μ^* are determined by condition that the resource constraint holds with equality.

³In deriving the optimal allocation, we again ignore the constraint that y is only observed by the firm. Again, we show below that this allocation can be implemented.

Condition (9) shows that marginal utility is equalized across employment and unemployment. Because $B > 0$ however, this implies that utility is lower when unemployed.

Condition (10) shows that the threshold level of productivity, \bar{y}^* , is lower than the production-efficient level b .⁴

2.2 Implementation

Assume, as before, that the state first chooses taxes and benefits, the firms then enter and offer a wage to workers, and, finally, productivity is realized. To describe how the optimal allocation is implemented, we again work backwards in time.⁵

At *Stage 3*, the threshold productivity below which the firm lays a worker off is still given by:

$$\bar{y} = w + \tau - f. \quad (3)$$

At *Stage 2*, the wage must satisfy the free entry condition:

$$V_F \equiv -G(\bar{y})f + \int_{\bar{y}}^1 y dG(y) - (1 - G(\bar{y}))(w + \tau) = I.$$

Consider thus the *problem faced by the government* in choosing taxes and unemployment benefits at *Stage 1*. From equations (3) and (10), it follows that, to induce firms to take the socially-optimal layoff decision $\bar{y}^* = b - B/U'(w)$, the following condition must hold:

$$f - \tau = \mu + \frac{B}{U'(w)}.$$

The net fiscal cost to the firm of laying off a worker must exceed the unemployment benefits paid to the worker by an amount which depends on the cost of becoming unemployed.

The other condition on taxes and benefits comes from the government budget constraint:

$$-G(\bar{y})(\mu - f) + (1 - G(\bar{y}))\tau = 0.$$

⁴This overemployment result should be familiar from the “implicit contract literature” (in particular Baily (1974), Azariadis (1975), Akerlof and Miyazaki (1980)): The lower layoff rate serves as a partial substitute for unemployment insurance.

⁵We continue to assume that severance payments are equal to zero. In this case, the assumption is not innocuous: Allowing for both unemployment benefits and severance payments would lead to a co-insurance problem and a suboptimal allocation. We discuss the issue in Section 6.

Combining these two conditions gives:

$$f = \mu + \frac{B}{U'(w)}, \quad \tau < 0 \quad (11)$$

The layoff tax must exceed unemployment benefits, implying, for budget balance, a negative payroll tax.

We summarize our results in Proposition 2.

Proposition 2. (i) *In the presence of limits to insurance, the threshold productivity in the socially efficient allocation is lower than the wage equivalent of being unemployed ($\bar{y}^* < b$): The lower incidence of unemployment partly compensates for the limits on insurance when unemployed.*

(ii) *Unemployment benefits, μ , must be financed by a combination of layoff taxes which exceed these benefits ($f > \mu$) and of negative payroll taxes, ($\tau < 0$). Put another way, the contribution rate must now be greater than one.*

2.3 Implications and discussion

- The main result in Proposition 2 is that, in the presence of limits to insurance, layoff taxes must exceed unemployment benefits — the contribution rate must be greater than one. The higher the utility cost of unemployment, the lower the layoff rate, the larger the layoff taxes.

This implies that, with respect to variations in B , unemployment insurance and employment protection are therefore substitutes: The lower the feasible level of insurance, the higher the implied level of employment protection. Higher employment protection however is a poor substitute for insurance, as it leads to distortions in the layoff decision. Thus, unemployment insurance reforms which reduce B (we shall give an example below, when interpreting B as a result of the need for search incentives) have both favorable direct (better insurance) and indirect (lower layoff taxes and lower distortions) effects.

- The second result is that payments by firms in case of layoffs must be larger than payments to the laid-off workers. This obviously could not be achieved by severance payments, which imply equal payments by firms and payments to workers, and requires the presence of a third party.⁶ We have taken this third party to be the state, collecting layoff taxes and

⁶This is an example of the general proposition (for example Holmström (1982)) that, when both incen-

(negative) payroll taxes, and paying unemployment benefits to workers. Formally, what is needed is a pooling or insurance agency, collecting payments from firms that layoff, paying unemployment benefits to workers, and distributing the difference to the remaining firms. In this case, firms who join are better off. The agency may therefore be private and participation voluntary.

- We formalized limits to insurance as coming from a non-financial cost of being unemployed. The limits may come instead from incentives.

Consider for example a modification of the benchmark based on shirking. Once hired, but before productivity is revealed, the worker decides whether to shirk or not. Shirking brings private benefits B but results in zero productivity and thus a layoff. Shirking is unobservable. Thus, to prevent shirking, the following condition must hold:

$$(1 - G(\bar{y}))(U(w) - U(b + \mu)) \geq B. \quad (12)$$

The expected utility gain from being employed relative to being unemployed must exceed some value B . The characterization of the optimum and the implementation are then identical to those above, except for the replacement of $B/U'(w)$ by $(1 - G(\bar{y}))B/U'(w)$ in the relevant equations.⁷

An alternative rationalization comes from the need to motivate the unemployed to search. While we cannot formally analyze this case in our one-period model, search incentive constraints are likely to lead however to results similar to those we have derived. The difference in utility between unemployment and employment has to be sufficient to induce search effort, yielding a constraint of the form $(U(w) - U(b + \mu) \geq B)$. A full treatment would however require a dynamic model, and we cannot provide it here.

3 Shallow pockets

In our benchmark, firms were risk neutral and had deep pockets. These assumptions are again too strong. Even in the absence of aggregate risk, the owners of many firms, especially small ones, are not fully diversified, and thus are likely to act as if they were risk averse.

tives and insurance considerations are present, there is typically a need for a “budget breaker”, such as an insurance company or the state.

⁷Under the more general assumption that shirking does not yield zero productivity but instead shifts the distribution of y from $G(\cdot)$ to $H(\cdot)$, with $G(\cdot)$ stochastically dominating $H(\cdot)$, results are less clear cut. In the absence of further restrictions, it is not necessarily the case that f is greater than μ .

And, even if entrepreneurs are risk neutral, information problems in financial markets are likely to lead to restrictions on the funds available to firms. In this section, we focus on the implications of limited funds.

We assume that each entrepreneur starts with assets $I + \bar{f}$, where $\bar{f} \geq 0$ is therefore the free cash flow available to the firm after investment. While it is a convenient starting point, the exogeneity of \bar{f} is obviously too stark, and we discuss a number of extensions below.

3.1 The optimal allocation

The government budget constraint (7), the threshold condition (3), and the condition that payments by the firm in case of layoff cannot exceed free cash flow ($f \leq \bar{f}$), can be combined to give the following constraint on \bar{y}, w and μ :^{8 9}

$$G(\bar{y})\mu - (1 - G(\bar{y}))(\bar{y} - w) \leq \bar{f}.$$

Therefore, the optimal allocation is the solution to:

$$\max_{\{w, \mu, \bar{y}\}} V_W \equiv G(\bar{y})U(b + \mu) + (1 - G(\bar{y}))U(w),$$

subject to the resource constraint:

$$V \equiv -G(\bar{y})\mu + \int_{\bar{y}}^1 y dG(y) - (1 - G(\bar{y}))w = I,$$

and the additional constraint:

$$G(\bar{y})\mu - (1 - G(\bar{y}))(\bar{y} - w) \leq \bar{f}.$$

⁸One may wonder whether allowing for job creation subsidies/taxes in addition to payroll and layoff taxes might alleviate the shallow pocket constraint, and improve the allocation. This is not the case. Subsidies, even if allowed in the government budget constraint, would not appear in the equation below.

⁹This constraint is derived as follows. First rewrite the threshold condition as $\tau = \bar{y} - w + f$ and replace τ in the government budget constraint to get $-G(\bar{y})(\mu - f) + (1 - G(\bar{y}))(\bar{y} - w + f) = 0$: For a given $\bar{y} - w$, the lower f , the lower is μ . Reorganize and use $f \leq \bar{f}$ to get the equation in the text.

From the first-order conditions, it follows that the worker still receives full insurance¹⁰:

$$w^* = b + \mu^*.$$

Furthermore, if the second constraint is binding (that is, if \bar{f} is less than unemployment benefits in the optimal allocation derived in Section 1) :

$$\bar{y}^* = b + \frac{(\mu^* - \bar{f})}{(1 - G(\bar{y}^*))} > b. \quad (13)$$

By limiting payments by firms in case of layoff, the shallow pocket constraint prevents the state from achieving the production-efficient threshold, and the layoff rate is now higher than the production-efficient level. The tighter the shallow pocket constraint — the lower \bar{f} , the larger $(\mu^* - \bar{f})$, the higher \bar{y}^* , and so, the larger the layoff rate. The levels of \bar{y}^* , w^* , and μ^* are determined by (13), the full insurance condition, and the condition that the resource constraint holds with equality.

The optimal allocation still offers full insurance to workers. But limited funds lead to a higher threshold productivity, and thus a higher layoff rate than in the benchmark.

3.2 Implementation

If the shallow-pocket constraint is binding, the state chooses the layoff tax $f = \bar{f}$. Given unemployment benefits μ^* , the government budget constraint then implies:

$$\tau = \frac{G(\bar{y}^*)}{1 - G(\bar{y}^*)}(\mu^* - \bar{f}) > 0.$$

As unemployment benefits exceed layoff taxes, payroll taxes must be positive.

The threshold productivity chosen by firms is therefore given by:

$$\bar{y}^* = b + \mu^* + \tau - f = b + \mu^* + \frac{G(\bar{y}^*)}{1 - G(\bar{y}^*)}(\mu^* - \bar{f}) - \bar{f} = b + \frac{(\mu^* - f)}{1 - G(\bar{y}^*)}.$$

¹⁰To see why the presence of shallow pockets does not prevent full insurance, consider an allocation where $w > b + \mu$. Now, consider a decrease in the wage of $\Delta w < 0$ and an equal increase in payments by firms to the state. This change affects neither the threshold condition nor the firm's profit. Use these increased payments to increase unemployment benefits by $-[(1 - G(\bar{y}))/G(\bar{y})]\Delta w$. Together, these changes imply a change in utility of $[-(1 - G(\bar{y}))U'(w) + (1 - G(\bar{y}))U'(b + \mu)](-\Delta w) > 0$. Thus, welfare can be improved until workers are fully insured.

This is the same expression as in (13), and so, layoff and payroll taxes indeed implement the optimal allocation. The derivation shows that we can think of the shallow pocket constraint as affecting the threshold productivity level directly (through the limit on the layoff tax) and indirectly (through the need for positive payroll taxes); both the lower layoff tax and the higher payroll tax reduce the fiscal cost of layoff for firms, and thus lead to a layoff rate higher than the production-efficient level.

By the same argument as before, the resource constraint implies that workers receive the optimal w^* and μ^* .

We summarize the results in Proposition 3.

Proposition 3. *In the presence of shallow pockets, workers remain fully insured ($w^* = b + \mu^*$). The threshold productivity is higher than the wage equivalent of being unemployed ($\bar{y} > b$), leading to a higher layoff rate than in the benchmark.*

This allocation can be implemented by the government choosing unemployment benefits μ^ , and financing them partly through layoff taxes \bar{f} , and partly through payroll taxes, $\tau > 0$. Put another way, the implementation implies now a contribution rate smaller than one.*

As before, implementation can be achieved by a pooling agency, receiving contributions $\bar{f}G(\bar{y}^*)$ from firms that layoff and contributions $\tau(1 - G(\bar{y}^*))$ from those that do not, and paying unemployment benefits $G(\bar{y}^*)\mu^*$ to laid-off workers. Firms have an incentive to join, and the agency may therefore be private.

3.3 Discussion

Two extensions of this analysis of shallow pockets are explored more formally in a companion paper (Blanchard-Tirole 2005). Here is a short summary of the conclusions of that paper:

(a) *Endogenization of limited pocket-depth.* There are at least two reasons why firms have shallow pockets. The first reason is that they may not want to have deep pockets even if they can. This arises when, in contrast to the maintained assumption of this paper, the government policy is set after rather than before firms invest. Suppose, for example, that the state cannot commit and sets (τ, f, μ) after firms have invested and hired workers, but before they learn the productivity of the match. In this case, firms will obviously choose to be “judgment proof”, i.e. to have no assets left in case of layoff, so $\bar{f} = 0$. The threshold is

then given by: $\bar{y}^* = b + \mu^*/(1 - G(\bar{y}^*))$. The high threshold, and by implication, the high layoff rate, reflects two distortions, one coming from zero layoff taxes and the other from positive payroll taxes.

Second, firms may have limited access to external finance due to a dearth of pledgeable income. When the entrepreneurs enjoy a rent $R > 0$ in case of continuation, investors cannot appropriate the full surplus from continuation, which leads to more layoffs than in the benchmark. We find however that, in this case, the state should still make the “investor-entrepreneur coalition” fully accountable for the cost of layoffs, so the optimal contribution rate remains equal to one.

(b) *Multi-activity firms.* A concern often expressed by policy makers is that, even if the firm can pay the layoff taxes, it may have to close otherwise healthy activities. The companion paper, in a two-activity context, investigates the possibility and implication of such “spillovers” or “snowball” effects. It shows that the state in general does not want to collect the fully collectable amount (\bar{f} in the notation of this section) in case of partial layoffs. If it did collect the full amount, the marginal layoff tax on further layoffs would be equal to zero (as the state cannot collect more than \bar{f} .) This would in turn provide excessive incentives for the firm to shut down entirely when having low productivity only in one activity.

4 Ex-post wage bargaining

Our benchmark embodied the assumption that wages were set ex-ante, i.e. at the time of hiring. This had the implication that, by offering unemployment insurance to risk averse workers, a firm could not only offer a lower wage, but actually lower its expected labor costs.

To some extent however, there is always some room for ex-post bargaining. When this is the case, a firm which has to pay a layoff tax if it lays a worker off is in a weaker bargaining position vis-à-vis that worker; a worker who will receive unemployment benefits if laid off is in a stronger position. The layoff tax, the severance payments, and the provision of unemployment benefits both lead to higher, not lower, wages, and thus increase labor costs.

In this section, we therefore modify our earlier assumption about wage setting, assume ex-post wage bargaining instead, and characterize the optimal allocation and its imple-

mentation under ex-post wage bargaining. In order to avoid the complexities attached to bargaining under incomplete information, we assume, in this section only, that both the firm and the worker observe productivity ex post. Thus “ex-post wage bargaining” includes the assumption of symmetric information between worker and firm ex post.

4.1 A formalization of wage bargaining

The following formalization captures the effects of benefits and taxes on ex-post wage determination in a simple way:

Assume wage setting now takes place after productivity is realized, and is the outcome of a two-stage game. In stage 1, the worker makes a wage offer to the firm. The firm can either accept the offer or turn it down. If it turns it down, the wage is set in stage 2, either by the worker with probability β , or by the firm with probability $1 - \beta$.

Under the assumption that the firm chooses the threshold level of productivity so as to maximize profit ex-post, in stage 2, the highest wage the firm will accept, and therefore the wage offered by the worker, is equal to $y - \tau + f$. The lowest wage the worker will accept, and therefore the wage offered by the firm, is equal to $b + \mu$. Thus, the expected wage in stage 2 is given by $\beta(y - \tau + f) + (1 - \beta)(b + \mu)$. This implies that, in stage 1, the worker will make the highest offer acceptable by the (risk neutral) firm, i.e. an offer of

$$w(y) = \beta(y - \tau + f) + (1 - \beta)(b + \mu).$$

The higher the layoff tax, or the lower the payroll tax, or the higher the unemployment benefits, the higher is the wage.

The threshold value for productivity, \bar{y} , is given by the condition that $w(\bar{y}) + \tau - \bar{y} = f$. Using the expression for the wage and rearranging:

$$\bar{y} = b + \mu + \tau - f. \tag{14}$$

Note that the threshold is privately efficient. Just as in the benchmark, the combination of $f = \mu$, $\tau = 0$ would deliver the production-efficient threshold, $\bar{y} = b$. But, as we shall see, other considerations are now relevant.

Expression (14) allows us to rewrite the wage schedule as:

$$w(y) = (b + \mu) + \beta(y - \bar{y}). \tag{15}$$

The wage paid to the marginal worker, the worker in a job with productivity equal to the threshold level, is equal to $(b + \mu)$, the wage equivalent of being unemployed plus unemployment benefits and severance payments. The wage then increases with β times the difference between productivity and threshold productivity.

4.2 The optimal allocation

The optimal allocation solves the same problem as in the benchmark, subject to the additional constraint that the wage is no longer a decision variable, but is instead given by equation (15):

$$\max_{\{\mu, \bar{y}\}} G(\bar{y}) U(b + \mu) + \int_{\bar{y}}^1 U(w(y)) dG(y),$$

subject to the resource constraint:

$$-G(\bar{y})\mu + \int_{\bar{y}}^1 (y - w(y)) dG(y) = I,$$

and the wage relation

$$w(y) = b + \mu + \beta(y - \bar{y}).$$

Under pure ex-post wage bargaining, the scope for unemployment benefits to provide insurance to workers is extremely limited: An increase in unemployment benefits does not change the slope of the income schedule, increasing all wages by an amount equal to unemployment benefits. Put differently, the only degree of freedom is the location of the kink \bar{y} in the worker's income schedule. The solution can be characterized as follows:

The threshold level of productivity is implicitly defined by:

$$\bar{y}^* = b + \frac{\beta(1 - G(\bar{y}^*))}{g(\bar{y}^*)} \left[1 - \frac{E[U'|y \geq \bar{y}^*]}{E U'} \right], \quad (16)$$

where $E[U'|y \geq \bar{y}] \equiv (\int_{\bar{y}}^1 U'(w(y))dG(y))/(1 - G(\bar{y}))$ is the expected value of marginal utility if employed, and $E U' \equiv G(\bar{y}^*)U'(b + \mu) + \int_{\bar{y}^*}^1 U'(w(y)) dG(y)$ is the unconditional expected value of marginal utility.

So long as β is strictly positive, and workers strictly risk averse, then the expected marginal utility if employed is less than the unconditional expected marginal utility, and so \bar{y}^* is greater than b . The layoff rate exceeds the production-efficient level. Setting $\bar{y}^* = b$

(and $f = \mu$) would achieve production efficiency. Our formalization shows that it is optimal to choose a threshold higher than the production-efficient level so as to decrease income uncertainty. The more risk averse the workers, or the stronger the workers in bargaining, the higher the threshold level, and so the higher the layoff rate.¹¹

Given \bar{y} , μ and by implication the wage schedule $w(y)$ are determined by the resource constraint above.

4.3 Implementation through payroll and layoff taxes

Replacing \bar{y} from (16) in the expression for the threshold decision of firms, (14), gives the first relation between f and τ :

$$f - \tau = \mu - \beta \frac{(1 - G(\bar{y}^*))}{g(\bar{y}^*)} \left[1 - \frac{E[U'|y \geq \bar{y}^*]}{E U'} \right]. \quad (17)$$

The other relation is given, as before, by the budget constraint, equation (7). As the second term on the right side of equation (17) is now negative, $f - \tau < \mu$. Together with the government budget constraint, this implies $f < \mu$ and so, a contribution rate below one. The reason is clear from above: The optimal threshold is higher than the production-efficient level. This is achieved by lowering the contribution rate from its benchmark value, namely unity. By implication, payroll taxes must be positive, in order to finance the shortfall of the unemployment insurance system.

¹¹One way of getting more intuition for the optimal threshold is to combine the first and the second constraints, and reorganize to read:

$$G(\bar{y})(b + \mu) + \int_{\bar{y}}^1 ((b + \mu) + \beta(y - \bar{y})) dG(y) \leq G(\bar{y})b + \int_{\bar{y}}^1 y dG(y) - I.$$

The left side gives the value of total workers' income (the sum of wages, reservation wages, and benefits) in the economy. The right side gives the value of total income in the economy. The two are the same, as free entry implies zero profit income.

Suppose, counterfactually, that changes in \bar{y} do not affect total income, and so do not affect total workers' income. Then, changes in \bar{y} change the shape of the income profile without affecting the total workers' income. An increase in \bar{y} implies that more workers receive $(\mu + b)$. The overall workers' income remains the same, but the income schedule is flatter. Indeed, the best value of \bar{y} from the point of insurance is $\bar{y} = 1$, where all workers receive the same income.

The assumption that changes in \bar{y} do not affect total income is, however, clearly incorrect. Starting from $\bar{y} = b$, small changes in \bar{y} do not affect total income; this implies that, as it decreases risk, at least a small increase in \bar{y} is desirable; this explains why the optimal \bar{y} exceeds b . But, as \bar{y} increases, the output loss increases, and total income decreases. For $\bar{y} = 1$ for example, total income is just b and so μ must be equal to zero. This implies that the optimal value of \bar{y} is greater than b , but less than one.

Note that for $\beta = 0$, i.e. if workers have no bargaining power, then we obtain the same characterization as in the benchmark: $f = \mu$ and $\mu = w - b$. As β becomes positive, and the wage schedule is now increasing in productivity, it can be shown that the equations above imply:

$$\frac{df}{d\beta} < \frac{d\mu}{d\beta} < 0.$$

That is, both the unemployment benefit and the layoff tax decrease as the workers acquire more bargaining power, and the layoff tax falls faster, leading to a decreasing contribution rate.

We summarize our results as follows:

Proposition 4. *When wages are set through ex-post bargaining, utility for the marginal worker is the same as for the unemployed workers, and workers with higher productivity receive a higher wage. These outcomes are independent of the state's policy choices. It is optimal to induce a threshold higher than the production-efficient level: ($\bar{y}^* > b$). This choice decreases the uncertainty faced by workers at some cost in efficiency.*

This optimal allocation can in turn be implemented through a combination of layoff taxes, payroll taxes, and unemployment benefits. Layoff taxes must be less than unemployment benefits, with payroll taxes used to make up the difference; equivalently, the contribution rate must be less than one.

5 Heterogeneity

We have assumed so far that all workers and all firms were ex-ante identical. In reality, they clearly are not. Firms differ in the distribution of productivity shocks (or, more generally, the distribution of productivity and relative demand shocks) they face, and in their initial assets. Workers also differ in the distribution of productivity. We study in this section the implications of heterogeneity in productivity, both on the firm and on the worker side, both observed and unobserved.

5.1 Heterogeneity of firms

A worry often expressed by policy makers is that, if some firms have higher layoff rates than others, a layoff tax will penalize them more, and this may be undesirable.

To explore this idea, suppose there are two types of firms, “strong” and “weak”, which differ in their productivity distributions. The productivity of “strong firms” is drawn from cumulative distribution $G_H(\cdot)$ and that of “weak firms” from distribution $G_L(\cdot)$.¹² The distribution function of strong firms stochastically dominates that of weak firms: for all y in $(0, 1)$, $G_H(y) < G_L(y)$. The fraction of strong firms is equal to ρ .

Let us start with the assumption that this heterogeneity is unobserved, so the state is unable to tell the two types of firms apart. Let us also start with the assumption that the state sets a uniform policy (τ, f, μ) . We return to these two assumptions below.

Note that given the distribution assumptions and uniform taxation, strong firms have higher expected profits than weak firms. We assume that the number of strong firms is fixed. Thus, at the creation margin, the free entry condition is relevant for weak firms only. By contrast, both types may lay workers off, so the destruction margin is relevant for both types of firms.

The fact that the free entry condition is relevant only for weak firms makes it more difficult to follow the optimal allocation/implementation approach we have followed until now. We take instead the more “pedestrian” route of solving the optimization problem of the state given firms’ behavior.

We assume that the state maximizes the welfare of workers. Allowing the state to also put some weight on the positive rents earned by the owners of strong firms would not alter the results. The state’s optimization problem is given by:

$$\begin{aligned} \max_{\{\tau, f, \mu, w, \bar{y}\}} \quad & [\rho G_H(\bar{y}) + (1 - \rho) G_L(\bar{y})] U(b + \mu) \\ & + [\rho(1 - G_H(\bar{y})) + (1 - \rho)(1 - G_L(\bar{y}))] U(w), \end{aligned}$$

subject to the free-entry condition for weak firms

$$-G_L(\bar{y})f + \int_{\bar{y}}^1 (y - w - \tau) dG_L(y) = I,$$

the government budget constraint

$$\begin{aligned} & -[\rho G_H(\bar{y}) + (1 - \rho) G_L(\bar{y})] (\mu - f) \\ & + [\rho(1 - G_H(\bar{y})) + (1 - \rho)(1 - G_L(\bar{y}))] \tau = 0, \end{aligned}$$

¹²We assume that the weak firms are not too unproductive. Namely we assume that $bG_L(b) + \int_b^1 y dG_L(y) \geq I$: Weak firms have positive NPV for the production-efficient threshold.

and the threshold productivity condition, which is the same for weak and strong firms:

$$\bar{y} = w + \tau - f.$$

The solution can then be characterized as follows:

- The state fully insures workers: $w = b + \mu$.
- The threshold level of productivity is given by:

$$\bar{y} = b + \frac{\rho(G_L(\bar{y}) - G_H(\bar{y}))}{(\rho g_H(\bar{y}) + (1 - \rho)g_L(\bar{y}))}.$$

By the definition of weak and strong firms, $G_L(\bar{y}) > G_H(\bar{y})$. So, unless $\rho = 0$, the threshold level is higher than the efficient level.

This solution is in turn implemented by choosing layoff taxes lower than unemployment benefits, so through a contribution rate lower than unity. The difference is financed through payroll taxes.

The intuition for why the contribution rate is less than one is as follows: Both types of firms may lay workers off and so the destruction margin applies to all firms. By contrast, the creation margin corresponds to the weak firms only. The state can then improve workers' welfare by allowing for some cross-subsidy from strong to weak firms—transferring some of the rents of strong firms to wages. Because the state is unable to assess the firms' strength, the cross-subsidy operates through the contribution rate: Weak firms lay workers off more than strong firms and so benefit more from a contribution rate smaller than unity.

Can the state do better by offering *menus* and letting firms self select? Appendix 1 shows that while offering a menu improves efficiency, the solution carries the main characteristics of the uniform policy, namely cross-subsidization of weak firms by strong firms, with a contribution rate for weak firms below one.

If heterogeneity is observable, i.e. if the state is able to tell weak and strong firms apart, the state can do even better. Characterizing the optimal policy is straightforward: It is still optimal to subsidize weak firms. This can now be done however by offering a

job creation subsidy to the weak firms, while setting the net contribution rate equal to 1 for both types of firms. This achieves the desired redistribution from strong to weak firms, while avoiding distortions at the destruction margin. (Note that both layoff and payroll taxes must be higher than in the benchmark, as the extra revenue is needed to finance job creation subsidies to weak firms. The difference between the layoff tax and the payroll tax remains however equal to unemployment benefits, so there is no distortion at the destruction margin.)

We can summarize our results as follows:

Proposition 5. *Suppose that there are two types of firms: “strong” or “weak”. Strong firms have productivity distribution $G_H(\cdot)$, weak firms $G_L(\cdot)$, and $G_L(y) > G_H(y)$ for all y in $(0, 1)$.*

If this heterogeneity is unobserved, and if the state relies on a uniform policy, then the optimal policy is to fully insure workers ($w = b + \mu$), choose a threshold level higher than the efficient level ($\bar{y} > b$), and implement it through a contribution rate less than one.

If the state offers a menu instead, then the optimal menu separates the firms and has the following properties: Workers are fully insured. Strong firms face a net contribution rate of unity, and choose the efficient threshold. Weak firms face a net contribution rate below one, and the threshold exceeds the efficient level. The underlying mechanism remains the same, cross-subsidization of weak firms through the use of a lower contribution rate.

Whether the state offers menus or not, it is optimal, if the proportion of weak firms is small enough, to not have them operate at all—at the cost of some ex-ante unemployment—and to set a unit contribution rate.

If heterogeneity is observed by the state, then the optimal policy is to offer a job creation subsidy to weak firms, choose the efficient threshold for productivity, and implement it through a unit net contribution rate for both types of firms.

5.2 Heterogeneity of workers

Another frequently expressed worry is that some workers are more likely to be laid-off than others, and that a layoff tax may make firms more reluctant to hire them, and thus make these workers worse off.

To explore this idea, we set up a case very similar to that of firms. We assume there are two types of workers, “high-ability” and “low-ability” workers. High-ability workers

have a productivity distribution given by $G_H(\cdot)$, low-ability workers a distribution given by $G_L(\cdot)$, with $G_L(y) > G_H(y)$ for all y in $(0, 1)$.¹³ The fraction of workers with high ability is equal to ρ .

We assume that firms know the workers' abilities. Proceeding in parallel with the earlier case of heterogeneity in firms, we start by assuming both that the state does not know the workers' abilities, and that it chooses a uniform policy (τ, f, μ) .

We can simplify the set-up of the optimization problem, by noting that workers of each type will be fully insured, and that, because high-ability workers are more valuable to firms, they will be paid more, both when employed and when unemployed.

Suppose therefore that low-ability workers receive w when employed, and μ when unemployed, and are fully insured, so $w = b + \mu$. High-ability workers will then receive $w + \Delta$ when employed, and $\mu + \Delta$ when unemployed (so $w + \Delta = b + \mu + \Delta$), where Δ is the additional expected profit brought about to the firm by a high-ability worker. Noting that Δ does not affect the layoff decision, so the threshold productivity is the same for both types of workers, it follows that Δ is given by:

$$\Delta = [G_L(\bar{y}) - G_H(\bar{y})] f + \int_{\bar{y}}^1 [y - (b + \mu) - \tau] [dG_H(y) - dG_L(y)].$$

Payment of Δ can be achieved by wage-indexed unemployment benefits, so a worker who is paid $w + \Delta$ receives $\mu + \Delta$ (and the firm pays correspondingly higher layoff taxes when laying a high-ability worker off).

With this characterization of wage setting, and assuming that the state maximizes a utilitarian social welfare function, the optimal policy is the solution to:

$$\max_{\{\tau, f, \mu, \bar{y}\}} \{ (1 - \rho) U(b + \mu) + \rho U(b + \mu + \Delta) \},$$

subject to the free entry condition

$$-G_L(\bar{y}) f + \int_{\bar{y}}^1 [y - (b + \mu) - \tau] dG_L(y) = I,$$

¹³We assume that the low productivity workers are not too unproductive. Namely we assume that $bG_L(b) + \int_b^1 y dG_L(y) \geq I$: Under observed heterogeneity, the low-ability workers are sufficiently productive to justify investment by firms.

and the government budget constraint

$$[(1 - \rho) G_L(\bar{y}) + \rho G_H(\bar{y})] (f - \mu) + [(1 - \rho) [1 - G_L(\bar{y})] + \rho [1 - G_H(\bar{y})]] \tau = 0,$$

where Δ and \bar{y} have been defined above.

The solution has the following form:

- If the proportion of low-ability workers is high enough, then the threshold for productivity is given by:

$$\bar{y} = b + \frac{\rho(1 - \rho)[G_L(\bar{y}) - G_H(\bar{y})] [U'(\mu + b) - U'(\mu + b + \Delta)]}{[(1 - \rho)g_L(\bar{y}) + \rho g_H(\bar{y})] [(1 - \rho)U'(\mu + b) + \rho U'(\mu + b + \Delta)]}.$$

Note that $G_L(\bar{y}) - G_H(\bar{y})$ is positive, and so is $U'(\mu + b) - U'(\mu + b + \Delta)$. So the fraction on the right is positive, and the optimal threshold is higher than the efficient level.

This solution is in turn implemented by using a contribution rate below unity, with the rest of unemployment benefits being financed by payroll taxes.

- If the fraction of low-ability workers is small enough, however, then it is not optimal to facilitate their employability. No job is created for them, and they receive unemployment benefits μ for sure. In this case, threshold productivity is given by:

$$\bar{y} = b.$$

This solution is implemented by relying on a gross contribution rate greater than one: $f/\mu > 1$; and a net contribution rate equal to one: $f - \tau = \mu$. The payroll tax rate is in turn given by $\tau = [(1 - \rho)/\rho] \mu$.

The intuition for these results is as follows. When both types of workers are employed, the use of a contribution rate below one reduces the cost to firms of hiring a low-ability worker. This in turn leads to a higher relative wage for low-ability workers, and therefore reduces inequality, but at some cost in efficiency. If the proportion of low-ability workers is very low, it is more efficient to have them not work, and then to choose tax rates so as to have efficient separations for high-ability workers.

Because unemployment benefits for low-ability workers exceed corresponding layoff taxes, the solution requires having both a positive payroll tax and a correspondingly higher

layoff tax on high-ability workers. This increases revenues while maintaining a net contribution rate equal to one.¹⁴

Note that the logic of the results under firm and worker heterogeneity is similar: Weak workers (firms) are more likely to be laid off (to lay off), and an incomplete internalization of the externality of the layoff on the UI fund benefits the creation margin.

What if the state can offer menus and firms then self-select? It is straightforward to show that the state then offers two options, one aimed at firms that announce they have hired a high-ability worker, one aimed at firms that announce they have hired a low-ability worker. The first option has a net contribution rate equal to one. The second option has a net contribution rate below one. Both types of workers are fully insured. Thus, just as in the uniform case, heterogeneity leads to lower layoff taxes, but in this case only for low-ability workers.

What happens if differences between workers are observable? When the state can tell apart high-ability from low-ability workers—for example if workers are in bad health or disabled—then the optimal policy is to offer a job creation subsidy (or, equivalently, uniformly lower payroll and layoff taxes) to firms if they hire a low-ability worker. The net contribution rate can then be set equal to one for both types, and the destruction margin is undistorted for both types. In general, when differences are only partly observable, the solution combines job creation subsidies and a net contribution rate below one.

We summarize our results in the following proposition:

Proposition 6. *Suppose that there are two types of workers: “high-ability” and “low-ability”. High-ability workers have productivity distribution $G_H(\cdot)$, low-ability workers $G_L(\cdot)$, and $G_L(y) > G_H(y)$ for all y in $(0, 1)$.*

If this heterogeneity is unobserved and the state relies on a uniform policy, then the optimal policy is to fully insure workers ($w = b + \mu$), choose a threshold level higher than the efficient level ($\bar{y} > b$), and implement it through a contribution rate less than one. The lower layoff tax leads in effect to a transfer from high-ability workers to low-ability workers.

If the state offers a menu instead, then the optimal menu separates workers and has the following properties: Workers are fully insured. Firms hiring high-ability workers face

¹⁴This result is related to the result in Cahuc and Jolivet (2003) where the need to finance a public good also leads to higher layoff and payroll tax rates.

a net contribution rate of unity, and choose the efficient threshold. Firms hiring low-ability workers face a net contribution rate below one, and the threshold exceeds the efficient level. The underlying mechanism remains the same, cross-subsidization of low-ability workers through the use of a lower contribution rate.

Whether or not the state is allowed to use menus, it is optimal, if the proportion of low-ability workers is small enough, to leave them unemployed, and use a unit contribution rate for the remaining, high-ability workers.

If heterogeneity is observed by the state, then the optimal policy is to offer a job creation subsidy to low-ability workers, and choose an efficient threshold for productivity, and implement it through a unit contribution rate.

6 Mixing unemployment benefits and severance pay

We have assumed throughout that the state was the sole provider of insurance to workers. We saw how, even under this assumption, the state was able to implement the optimal allocation. This assumption however raises two issues relative to the scope for private-sector provision of insurance. The first is whether a private third-party insurer can substitute for the state. As we have seen, the answer is yes as long as redistributive concerns, across firms or across workers, are absent, i.e. except in section 5. The second issue is whether the third party insurer, whether the state or a private insurance company, needs to demand exclusivity and prohibit the firm from offering supplemental insurance in the form of severance pay.

A recurrent theme of the insurance literature is that the insurer must be wary of the externality imposed by supplemental insurance contracts (Pauly (1974)). For this reason, insurance companies often demand exclusivity and managerial compensation contracts prevent executives from undoing their incentives through insider trading or derivatives contracts with financial institutions. In the context of this paper, the state's provision of insurance to workers raises the question of welfare-reducing supplemental insurance.

Let us first note that allowing severance pay does not reduce (nor, of course, does it increase) welfare when the state already provides full insurance to workers. Intuition suggests, and analysis confirms, that the firm does not want to undo the full insurance provided by the state by overinsuring the worker (severance pay) or underinsuring her (asking the worker to return some of the unemployment benefits, assuming this were feasible). Thus, the analysis of Sections 1 (benchmark), 3 (shallow pockets) and 5 (firm or worker

heterogeneity) is immune to the introduction of severance pay.

The co-insurance problem may therefore arise only when the worker is imperfectly insured, that is in Sections 2 (incomplete insurance) and 4 (ex post wage bargaining). In Section 4, though, the firm is unable to exchange a bit more insurance against a lower wage. Indeed, severance pay increases the wage one-for-one; if μ_F denotes the severance pay, the new wage function, $\hat{w}(y)$, can be written as:

$$\begin{aligned}\hat{w}(y) &= \beta (y - \tau + f + \mu_F) + (1 - \beta) (b + \mu + \mu_F) \\ &= w(y) + \mu_F.\end{aligned}$$

The threshold value for productivity, \bar{y} , is given by the condition that $\hat{w}(\bar{y}) + \tau - \bar{y} = f + \mu_F$; that is, the threshold

$$\bar{y} = b + \mu + \tau - f$$

is independent of severance pay. Put differently, the introduction of severance pay exerts no externality on the state (as it doesn't change \bar{y}) and only serves to increase labor costs. Thus, the analysis of Section 4 is also immune to the introduction of severance pay.

The only case in which co-insurance reduces welfare and exclusivity is therefore optimal is that of incomplete insurance (Section 2). To see this, return to stage 2 and allow firms to offer, as in the benchmark, contracts which specify both a wage w and a severance payment, μ_F . The expected utility of workers is given by:

$$V_W \equiv G(\bar{y})(U(b + \mu_F + \mu) - B) + (1 - G(\bar{y}))U(w),$$

And the free entry condition is given by:

$$V_F \equiv -G(\bar{y})(f + \mu_F) + \int_{\bar{y}}^1 y dG(y) - (1 - G(\bar{y}))(w + \tau) = I.$$

Now, starting from $\mu_F = 0$, and assuming the economy is at the optimal allocation, so $w = b + \mu$, consider the effects of a small increase in μ_F . From the two equations above, it follows that

$$dV_W = \frac{g(\bar{y})}{1 - G(\bar{y})} B d\mu_F > 0.$$

Firms therefore have an incentive to offer more insurance than required in the optimal allocation. The reason why is that increasing μ_F has two effects on the expected utility of

workers. First, it creates a wedge between marginal utility when employed and unemployed; starting from the optimal allocation, this effect is of second order. The other is that it reduces the probability of a layoff; because the loss in utility from becoming unemployed is equal to $U(w) - U(b + \mu) + B = B$, this effect is of first order and dominates the first. When firms increase μ_F however, they decrease layoffs, and given that layoff taxes exceed unemployment benefits paid by the state, they impose a negative externality on the state. This is why, in the end, letting firms freely choose severance payments leads to a suboptimal allocation.

Proposition 7. *In the incomplete insurance extension of Section 2, allowing the firm to pay severance payments in addition to the unemployment benefits paid by the state leads to oversinsurance, an unemployment insurance deficit and suboptimal welfare. By contrast, severance pay has no impact on welfare in the benchmark and in the other extensions.*

7 Extensions and Conclusions

We are very much aware however of the limits of our analysis. Even within our one-period model, there are a number of issues still to be explored. Let us mention two.

An important issue not taken up here is that of quits versus layoffs. If we think of layoffs as triggered by productivity shocks (shocks to y), and quits as triggered by reservation wage shocks (shocks to b , or to the disutility of work—which we do not have explicitly in our model), and we think of the layoff tax as applying only in case of layoffs, this raises two sets of issues. The first is actions by firms to induce workers they would like to lay off to quit instead (harassment), and actions by workers to induce firms they would like to quit to lay them off instead (shirking). The second is actions by firms and workers together to mislabel quits and layoffs. The incentives to harass, shirk, or cooperatively misreport, depend very much in each case on the contribution rate. We have informally explored these issues in Blanchard and Tirole (2003b), but a formal treatment remains to be given.

Another issue is the role of judges, who, in many European countries, play a central role, and are often ultimately in charge of deciding whether layoffs are justified or not. Clearly, the logic of our argument is that this is better accomplished through a combination of layoff taxes and severance payments, with the decision then being left to the firm. But our look at the implications of imperfections, from shallow pockets to heterogeneity, also suggests the

desirability of adapting layoff taxes to particular situations. This can in principle be done through offering menus, or allowing taxes to be conditional on observable characteristics of firms, or by leaving some discretion to judges. It remains to be shown however if and when judges do in fact have the information, the ability, and the incentives, to take better and more informed decisions.

Then, and obviously so, there are dynamic issues we could not consider at all in our one-period model. Dynamic models of the labor market with risk aversion and imperfections are notoriously hard to solve. The only model we know which derives optimal institutions—defined as the optimal combination of payroll taxes, layoff taxes, job creation subsidies or taxes, and unemployment benefits—was developed by Mortensen and Pissarides (2003). However, it assumes risk neutrality and so cannot deal in a convincing way with the interaction between insurance and efficiency. (A model by Alvarez and Veracierto [1998] has risk averse workers, self-insurance as well as state-provided insurance, payroll and layoff taxes, and severance payments. While it shows (numerically) the effects of changes in some of these instruments, it does not give a characterization of optimal taxes and benefits.) We see two extensions of our model as essential:

The first is the role and the implications of self insurance by workers (in terms of the model here, the role and implications of the endogeneity of b). In this context, an important question is the role, if any, of mandatory individual unemployment accounts such as are being considered or introduced in a number of Latin American countries. (Three papers provide a useful starting point here. All three allow for self-insurance, and look at the role of state-provided insurance in the presence of other imperfections. In Hansen and Imrohoroglu (1992), moral hazard in search limits the scope for state-provided insurance. In Acemoglu and Shimer (1999, 2000), state-provided insurance affects search, which in turn affects match quality.)

The second is the role of experience rating systems, such as the US system, as ways of implementing the collection of layoff taxes. This requires a careful look not only at the dynamic problem of the firm, but at the exact nature of the financial constraints that it faces.

We feel, nevertheless, that as it stands, the paper can be helpful in thinking about reform about employment protection in Europe (the origin of the paper was indeed a request to define the contours of employment protection reform in France; our conclusions are presented in Blanchard and Tirole [2003a,b]):

The basic conclusion from the benchmark that unemployment benefits should be financed through layoff taxes is straightforward conceptually. It is however very much at odds with reality in Europe. All European unemployment insurance systems are financed through payroll rather than layoff taxes; other things equal, both lead to too high a layoff rate.

The lack of layoff taxes is counterbalanced however, in many countries, by heavy judicial intervention. This suggests that a shift from judicial intervention to higher layoff taxes on firms may be desirable, reducing uncertainty and channeling layoff costs into unemployment benefits.

The basic conclusion from considering limits to insurance—that in the presence of limits to insurance, higher layoff taxes and a contribution rate larger than one are justified—is also very relevant. Employment protection is then a partial substitute for insurance, but this comes at the cost of production efficiency. This suggests the importance of reforms of unemployment insurance which allow for better insurance while still providing incentives to search. Recent reforms, which make unemployment benefits more explicitly conditional on search and acceptance of jobs if available, go in that direction. If successful, they can bring not only better insurance, but also lower employment protection and lower production inefficiencies.

The final broad insight is that layoff and payroll tax schedules should be tailored to the economic environment in which firms operate. Lower layoff taxes and higher payroll taxes are called for in industries, countries, or periods in which firms have shallow pockets. Similarly, unobserved firm or worker heterogeneity suggests relying more on payroll taxes, and less on layoff taxes in the financing of unemployment benefits.

References

- [1] Acemoglu, D., and R. Shimer, (1999), “Efficient Unemployment Insurance”, *Journal of Political Economy*, 107-5, October, 893-928.
- [2] Acemoglu, D., and R. Shimer, (2000), “Productivity Gains from Unemployment Insurance”, *European Economic Review*, 44-7, June, 1195-1224.
- [3] Akerlof, G., and H. Miyazaki (1980) “The Implicit Contract Theory of Unemployment Meets the Wage Bill Argument,” *Review of Economic Studies*, 48: 321–338.
- [4] Alvarez, F., and M. Veracierto (1998), “Search, Self-Insurance and Job Security provisions,” working paper WP98-2, Federal Reserve Bank of Chicago.
- [5] Azariadis, C. (1975) “Implicit Contracts and Underemployment Equilibria,” *Journal of Political Economy*, 83: 1183–1202.
- [6] Baily, M.(1974) “Wages and Employment under Uncertain Demand,“ *Review of Economic Studies*, 41: 37–50.
- [7] Blanchard, O., and J. Tirole (2003a), “Redesigning the Employment Protection System,” *De Economist*, 152: 1-20.
- [8] — (2003b), “Licenciements et Institutions du Marché du Travail,” Rapport pour le Conseil d’Analyse Economique. La Documentation Française, pp. 7–50.
- [9] — (2005), “Layoff Taxes and Shallow Pocket Firms,” mimeo, MIT and University of Toulouse.
- [10] Cahuc, P. and G. Jolivet, (2003), “Do We Need More Stringent Employment Protection Legislation?”, mimeo Paris I.
- [11] Hansen, G. and A. Imrohoroglu, (1992) “The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard”, *Journal of Political Economy*, 100-1, February, 118-142.
- [12] Holmström, B. (1982) “Moral Hazard in Teams,” *Bell Journal of Economics*, 13-2: 324-340.

- [13] Mortensen, D., and C. Pissarides (2003) “Taxes, Subsidies and Equilibrium Labor Market Outcomes”, *Designing Inclusion: Tools to Raise Low-End Pay and Employment in Private Enterprise*, *Edmund S. Phelps (ed)*, Cambridge: Cambridge University Press.
- [14] Pauly, M. (1974) “Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection”, *Quarterly Journal of Economics*, 88(1): 44-62.

Appendix 1: Policy menus under firm heterogeneity

Suppose the state offers an option (τ_L, f_L) targeted at weak firms and another (τ_H, f_H) at strong firms, rather than the single option (τ, f) . With obvious notation change (thresholds, payroll and layoff taxes are now indexed by the type of firm), the optimization problem is identical to the problem above, except that the last constraint,

$$\bar{y} - [w + \tau - f] = 0,$$

is replaced by the constraints:

$$\bar{y}_L - [w + \tau_L - f_L] = 0, \quad \bar{y}_H - [w + \tau_H - f_H] = 0,$$

and the incentive compatibility constraint that the strong firms do not want to masquerade as weak ones is given by:

$$-G_H(\bar{y}_H) f_H + \int_{\bar{y}_H}^1 (y - w - \tau_H) dG_H(y) \geq -G_H(\bar{y}_L) f_L + \int_{\bar{y}_L}^1 (y - w - \tau_L) dG_H(y).$$

It is useful at this point to define the *net contribution rate* as the ratio of the layoff tax minus the payroll tax to unemployment benefits, $(f_i - \tau_i)/\mu_i$, $i = H, L$. Note that, so long as the net contribution rate is equal to one, the layoff decision is efficient: What matters here is not the layoff tax itself, but the difference between the layoff tax and the payroll tax.

The solution can then be characterized as follows:

- The state still fully insures workers: $w = b + \mu$.
- Strong firms face a net contribution rate equal to one, and so choose the efficient threshold:

$$\bar{y}_H = b \text{ and } f_H - \tau_H = \mu.$$

- Weak firms face a net contribution rate below one, and so choose a threshold higher than the efficient level:

$$\bar{y}_L = b + \rho \frac{G_L(\bar{y}_L) - G_H(\bar{y}_L)}{(1 - \rho)g_L(\bar{y}_L)} \text{ and } f_L - \tau_L \leq \mu.$$